Nature of Cardiovascular diseases

By

Mawuli Agboklu

EXECUTIVE SUMMARY

This is an analysis of understanding the nature of cardiovascular diseases with the data from the Framingham heart study. The study is a continuous study of male and females from Framingham, Massachusetts who had not yet developed overt symptoms of cardiovascular disease or suffered a heart attack or stroke.

I performed some data analysis on the data by developing 5 patient profiles, filtered and clustered them to observe patterns amongst them. I also provided recommendations on what I think should be done to avert death and extreme cases.

In performing the analysis on the Framingham dataset, filtering was the main data preparation technique used to generate/create clusters for the 5 patient profiles for further in-depth analysis.

To begin with, the dataset has 5209 observations with 17 variables some relating to each other and others not. This is a lot of data to interpret and make sense of also, some of the variables Ire not important to our task hence there was the need to filter out some variables from the set and split the observations into different clusters to allow us have clarity and focus on what matters. I created 5 filtered profiles namely filter female, filter male, filter alive, filter dead and filter cause of death. Each of these filters has 5 clusters and segment profile which profiles each cluster and compares their results in relation to the original set of data.

FILTER FEMALE

The Filter female profile is a filter of only females from the Framingham dataset. This was created to observe the different trends amongst only females. This filter had 2756 observations which were only females to be trained. We then created a cluster of 5 segments to understand the data some more. Segment 1 has a frequency of 741 segment 2 had 784, segment 3 had 686, segment 4 had 110 and segment 5 had 435.

Segment 1 has a different behavior. The top 3 variable worth here are Age-At-Start, MRW, and Weight Status. Most participants in the study in this segment started long after 20 years old whereas comparing it to the entire population, most started before age 20. This segment has very high MRW just as the entire population which means the average female from this segment is overweight. This translates to 93% of the segment population being recorded as overweight with an almost insignificant 6% recorded as having normal weight status which agrees with the pattern of figures for the entire population of 65% overweight, 30% normal weight and 4% underweight.

Segment 2. The top 3 variable worth for this profile are MRW (Metropolitan Relative Weight), Weight Status and BP (Blood Pressure) status. From the segment analysis, segment 2 showed that most of the females have lower weights on the average as compared to the entire population which has higher MRWs. This resulted in a positive weight status of the segment with 76% of the cluster segment recording normal weight status and 23% recording overweight. With the blood pressure variable, the

segment records 53% of the segment population as having optimal status, 41% as normal and 5% as high status. This slightly contrasts with the entire population which has 42% as normal status 39 % as high status and 18 % as having optimal blood pressure status. The increase in the entire population's high blood pressure status can be attributed to their high overweight numbers.

Segment 3 has systolic, blood pressure status and diastolic as its top 3 variables. Systolic is the upper number of a blood pressure measurement which means the force produced by the heart as it pumps blood out to the rest of the body. And diastolic represents the bottom number which means the pressure in your blood vessels when the heart is at rest. Doctors monitor the systolic especially because the higher it is, the greater the risk of a heart disease or stoke in the patient. In this segment, we see the systolic numbers in the segment population abnormally high. Higher as compared to the entire populations. This means the average female from this segment has a diagnosis of a high blood pressure. This has led to a 100% recording of high blood pressure amongst the segment population as well as high diastolic numbers.

Segment 5. This segment has the start age for most of its females far above that of the entire population. Majority of the segment population have a normal blood pressure status with fractions having high and optimal blood pressure statuses.
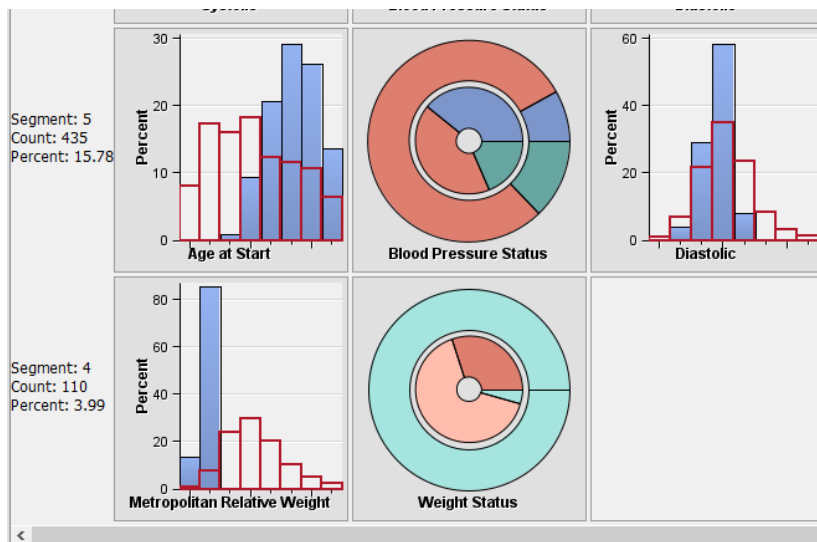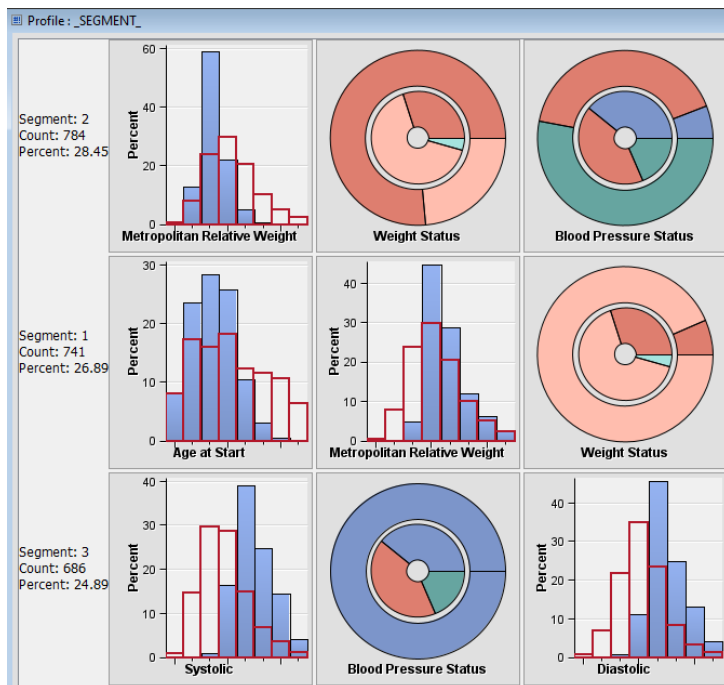
Figure 1- Train data of Female

FILTER ALIVE

This filter had 3172 observations which were only living participants trained. We then created a cluster of 5 segments to understand the data some more and see what the pattern was amongst living participants. Segment 1 has a frequency of 550 segment 2 had 672, segment 3 had 745, segment 4 had 458 and segment 5 had 747.

Segment 1 focused on the blood pressure status of its participants. The top 3 variables were systolic, blood pressure status and diastolic. Systolic which determines where one is hypertensive or not his extremely high amongst the segment population which had led to a whopping 97 % of them having very high-pressure statuses. However, a fraction 2% of them were able to maintain normal blood pressures. Once the systolic was high, chances that the diastolic would also be high in unavoidable.

Segment 2 also had similar variable worths like segment 3 however, it was discovered that, despite their high MRWs most of them had normal weights with just a fraction of them overweight and a handful being underweight. Most of the participants here were females.

Segment 3 had majority of its participants as females with 98% and 2% as males. The top 3 variables here were Sex, MRW and Weight Status. It was further discovered that they had very high MRWs which resulted in a 100% overweight status of the segment.

Segment 4. This final segment also focused on 3 key variables. Blood pressure status, diastolic and Sex. It was discovered that majority of the population had high blood pressure with high diastolic measurements and most of the participants in this segment were males with very few being females.

Segment 5 was entirely made up of male participants. The top 3 variables or this segment was sex, height, and blood pressure. The participants had heights higher than the original dataset when they were compared. They also had normal blood pressure statuses as much as 75%, optimal of 23% and a minute 1% of them had very high-pressure statuses.
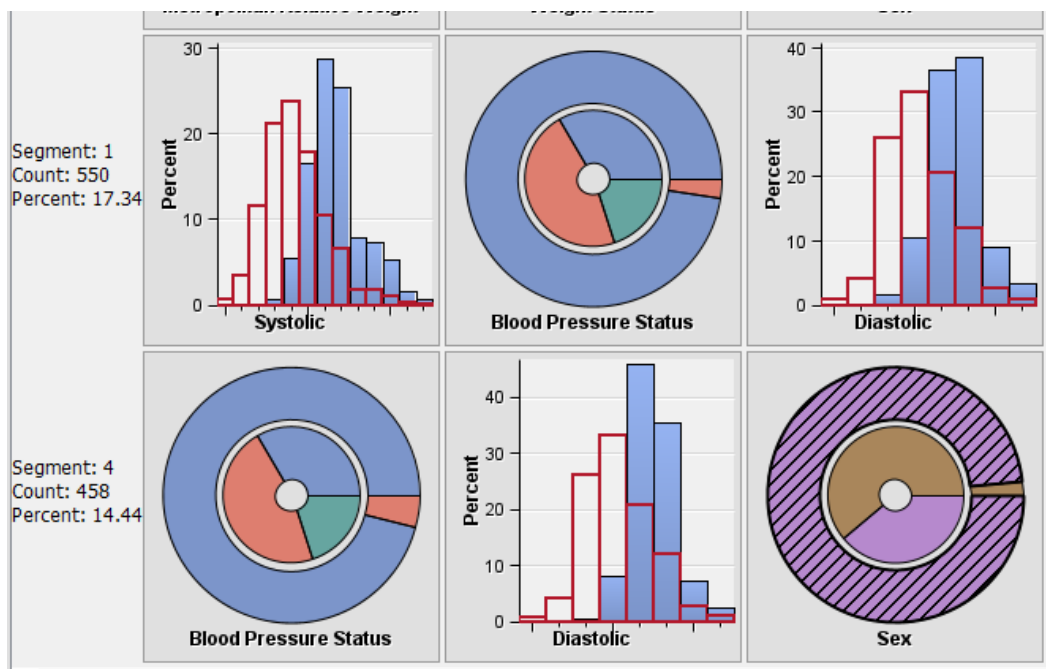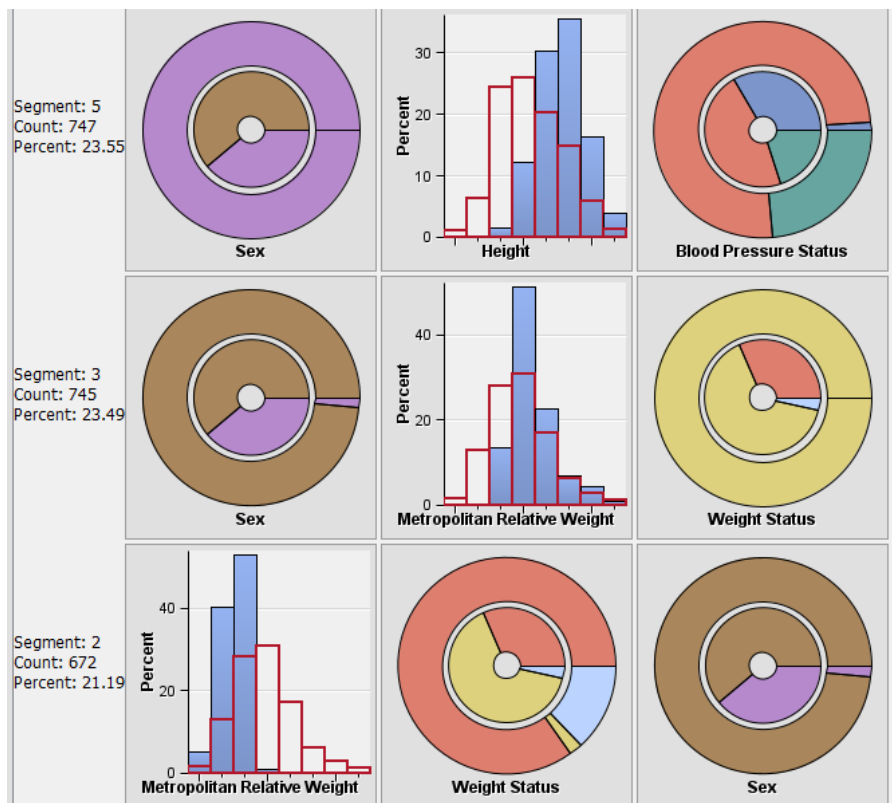
Segment: 5
Count: 747
Percent: 23.55

Segment: 3
Count: 745
Percent: 23.49

Segment: 2
Count: 672
Percent: 21.19

Segment: 1
Count: 550
Percent: 17.34

Segment: 4
Count: 458
Percent: 14.44

Figure 2- Train data of Alive

FILTER DEAD

This filter had 1878 observations which were only dead participants trained. We then created a cluster of 5 segments to understand the data some more and see what the pattern was amongst dead participants. Segment 1 has a frequency of 442 segment 2 had 322, segment 3 had 442, segment 4 had 206 and segment 5 had 495.

Segment 1. This segment samples were all males. Their diastolic averages were low, and their blood pressure levels were also mostly normal. It was just a handful of 17% that recorded high blood pressure and 14% recorded optimal blood pressure.

Segment 2. The start ages of the participants in this segment were way above the averages of the entire dataset population. The same applied to their ages at death. However, their systolic figures were moderate and not so high.

Segment 3. This segment was focused on sex, systolic and height. Most of the those sampled in this segment were 90% females and 10% male. They had very high systolic numbers which the data proved to have resulted in high blood pressure and their heights were also above the entire population samples.

Segment 4. The top 3 variable worth for this segment are sex, blood pressure status and height. Most of the considered participants were female, most of whom had normal blood pressures with a handful having high and optimal blood pressures. Their heights were above the dataset's average.

Segment 5. The top 3 variable worth in this category is diastolic, blood pressure status and sex. The participants in this category who happen to be dead had high diastolic numbers higher than the entire population, which caused high blood pressure amongst them. 95% of them were males and 5% females.
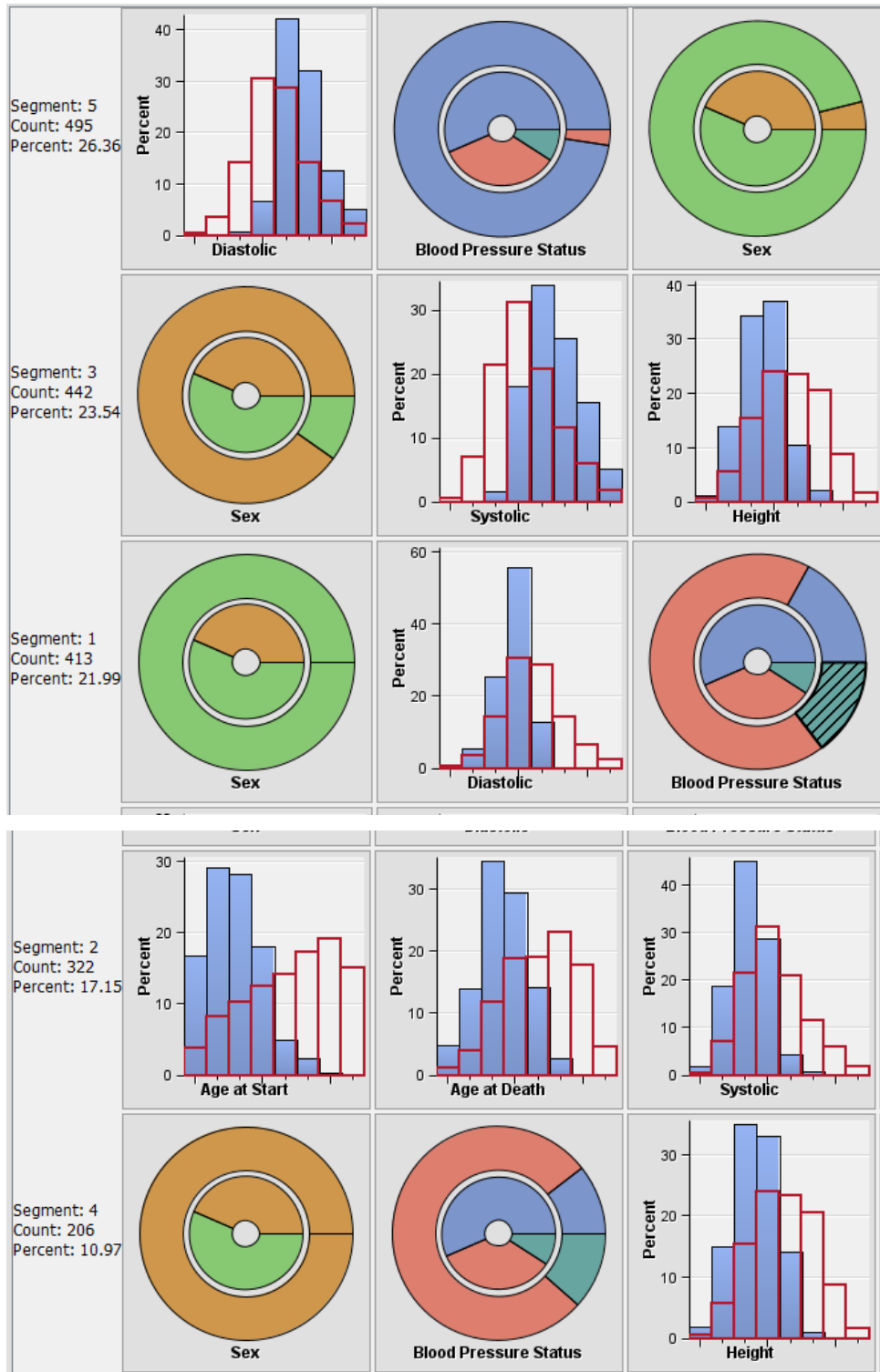
Figure 3- Train data of Dead

FILTER MALES

This filter has 2,336 observations which are only males from the Framingham dataset. This was created to observe the different trends among only males. The data was trained. I created a cluster of 5 segments to separate and understand the data. Segment 1 had a frequency of 596, segment 2 had a frequency of 65, segment 3 had a frequency of 418, segment 4 had a frequency of 759, and segment 5 had a frequency of 456.

Segment 1 The top 3 variables are diastolic, systolic, and blood pressure status. Men are above the national average for diastolic and systolic. Blood pressure status is high at 45.51% for the male population.

Segment 2 The variables are metropolitan weight and Wight status. Compared to the national average, males are above by 76.92%. Weight status of men are 70% overweight and 27.07% normal Weight.

Segment 3 The main variables are age at death, age at start, and cause of death. Men are above average in their later years of life. Men are below average in their earlier years. As they aged then they are double the national average. The main cause of death is coronary heart disease at 16.56% and running up is cancer at 12.20%.

Segment 4 has age at death, cause of death, and status as its top 3 variables. Men are below average for the cause of death. The cause of death in men was coronary heart disease at 16.56% and cancers at 12.21%. The status was 53.75% alive compared to 46.25% dead.

Segment 5 Men are double the national average on metropolitan relative weight at the highest being 51.31%. Men are overweight by 70% as compared to normal weight at 27.07% on the national population scale.
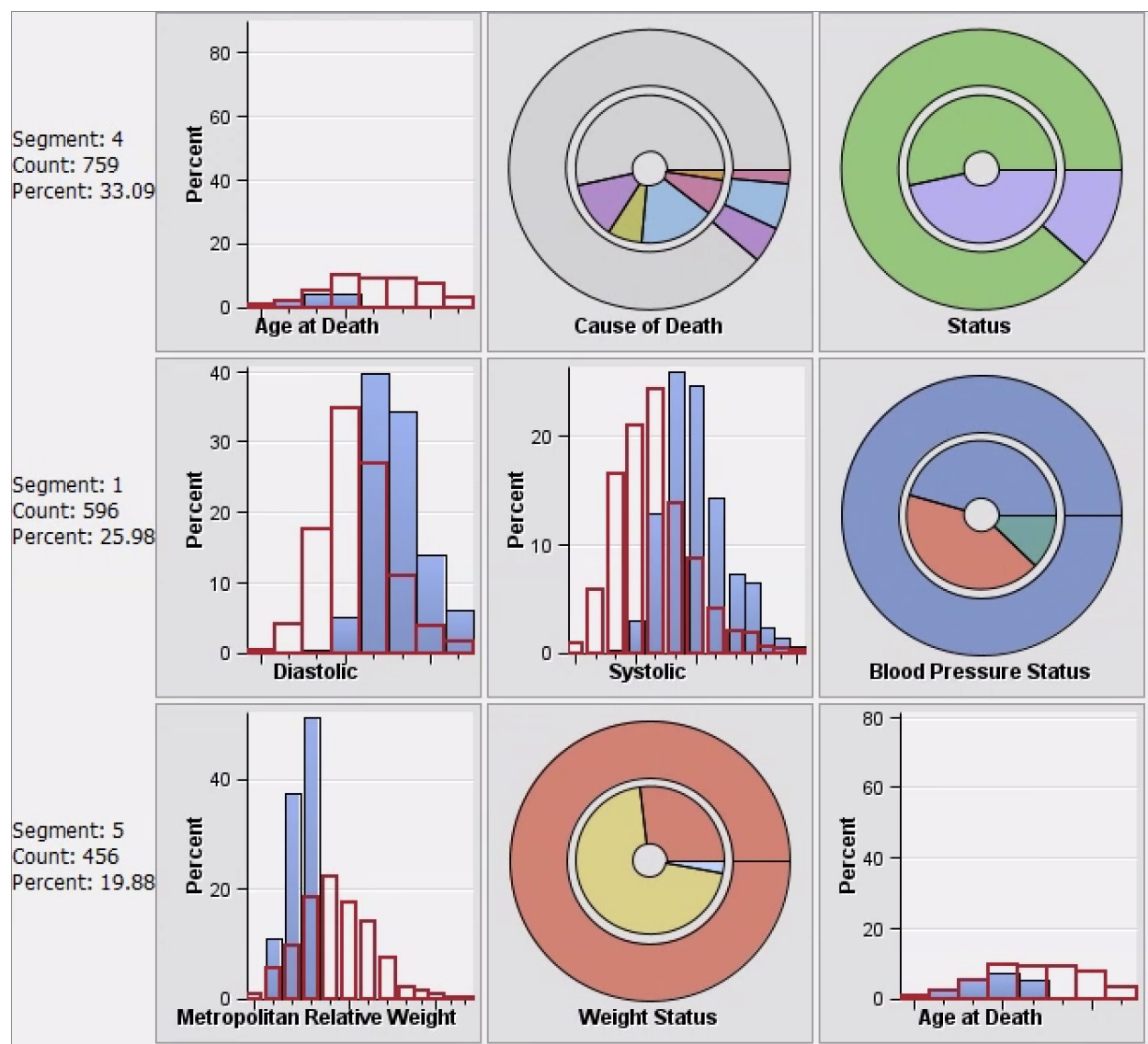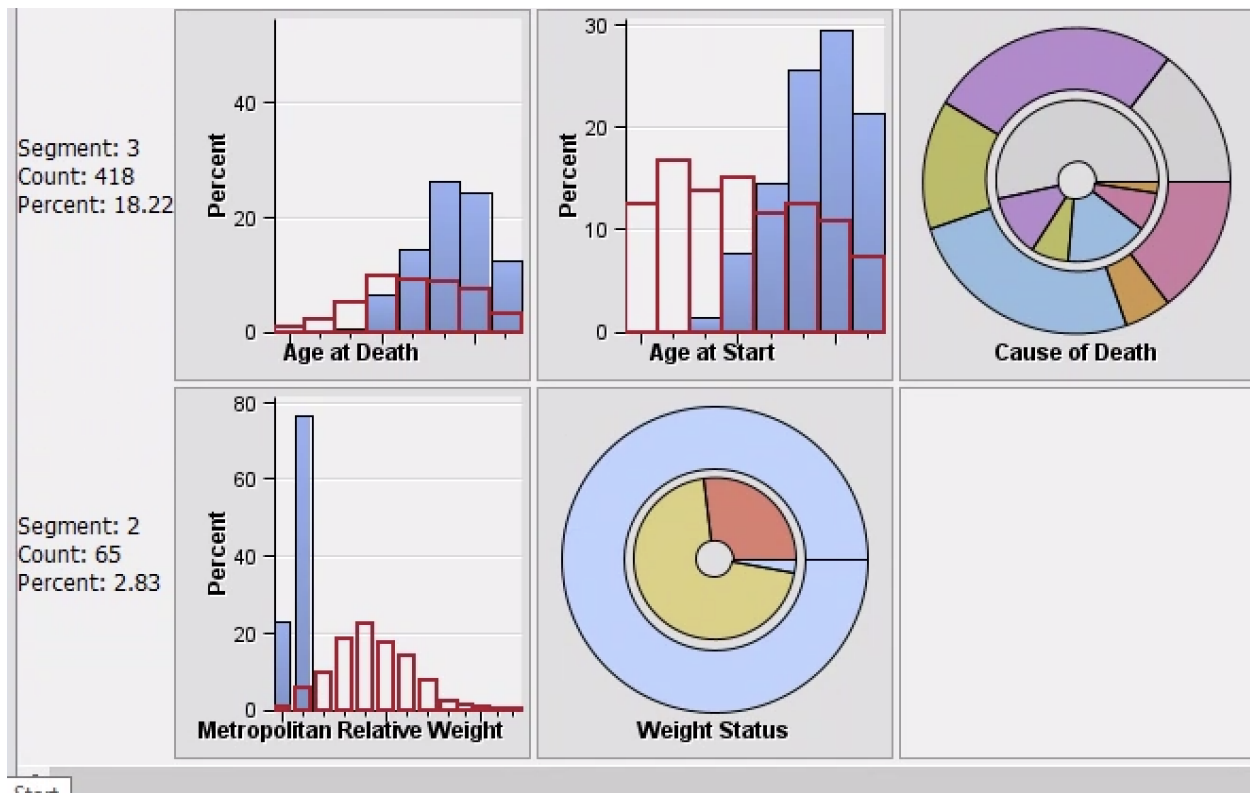
Figure 4- Train data of Men

<u>FILTER CAUSE OF DEATH</u>

This filter has 5,050 observations which both female and male cause of death was filtered. I then created a cluster of 5 segments to understand the data. Segment 1 has a frequency of 995, segment 2 has a frequency of 881, segment 3 has a frequency of 958, segment 4 had a frequency of 976, and segment 5 had a frequency of 1,240.

Segment 1 Systolic is above average with the highest at 40.30% for the cause of death in females and males. Blood pressure status for females and males is 42.21% normal, 42% high, and 15.78% optimal. Diastolic is above average with the highest at 45.12% for the cause of death in females and males.

Segment 2 Blood pressure status for females and males is 42.21% normal, 42% high, and 15.78% optimal. Diastolic is above average with the highest at 52.66% for the cause of death in females and males. Sex of females 54.57% and males 45.42% being the cause of death.

Segment 3 Metropolitan relative weight is above average for females and males at 61.06%. weight status for females and males is overweight at 67.48%, normal at 28.85%, and underweight at 3.54%. Systolic is above average with the highest at 40.08% for the cause of death in females and males.
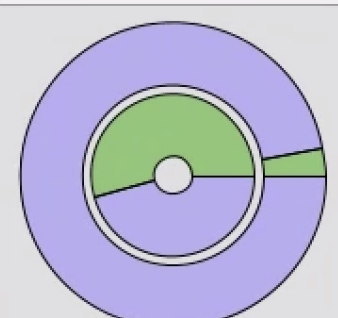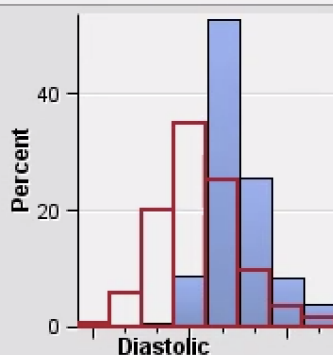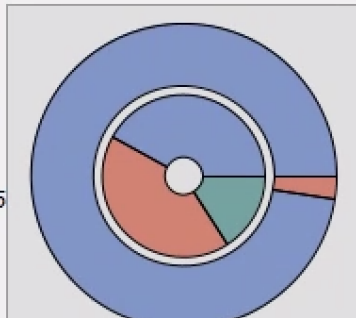
Segment 4 Sex of females 54.57% and males 45.42% being the cause of death. Height is above average with the highest at 43.43% for males and females. Metropolitan relative weight is above average for females and males at 49.69%.

Segment 5 The top 3 variables are sex, height, and blood pressure status. Cause of death for females 54.57 % and males 45.42%. Height for both females and males is above the national average with the top being 35.48%. Blood pressure status for females and males is 42.21% normal, 42% high, and 15.78% optimal.
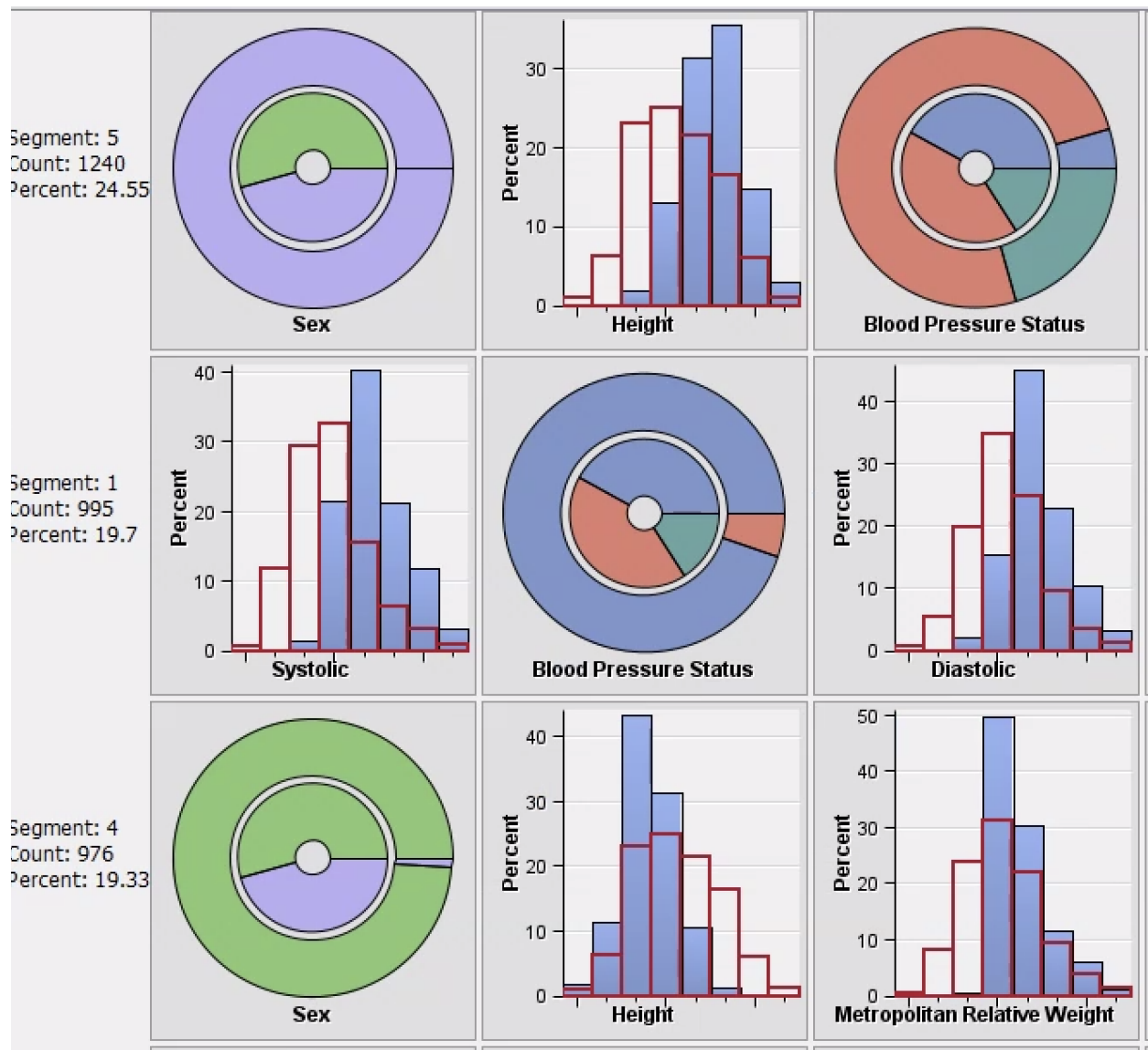
Segment: 3
Count: 958
Percent: 18.97

Metropolitan Relative Weight
Weight Status
Systolic
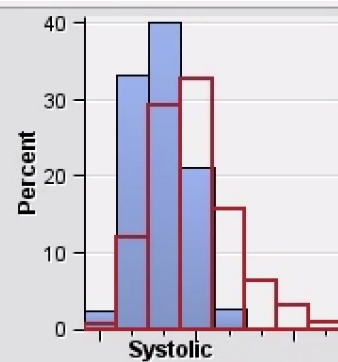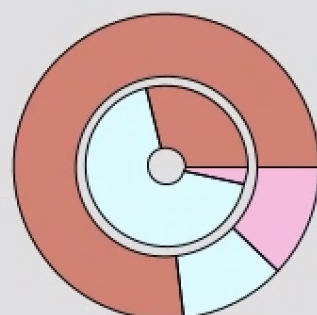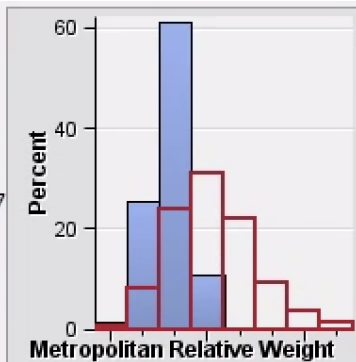
Segment: 2
Count: 881
Percent: 17.45

Blood Pressure Status
Diastolic
Sex

Figure 5- Train data Cause of Death

Death rate by Blood Pressure level & Sex

Optimal_Female 4.3%
Normal_Female 13.0%
High_Male 31.1%
High_Female 27.8%
Normal_Male 19.5%
Optimal_Male 4.3%
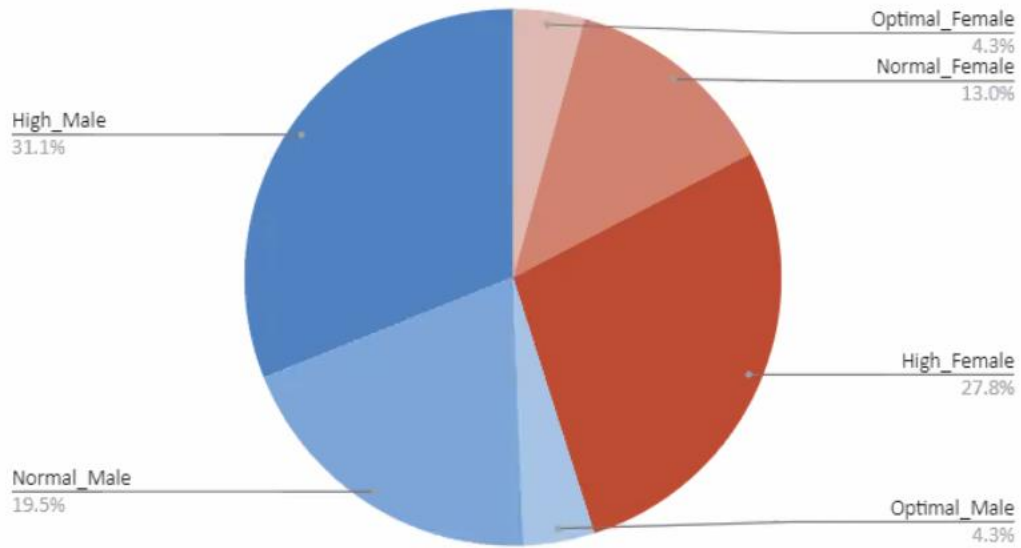
Figure 6- Death rate by Blood Pressure level and Sex

Death rate by Weight & Sex

Underweight_Female 1.7%
Normal_Female 10.1%
Overweight_Male 40.2%
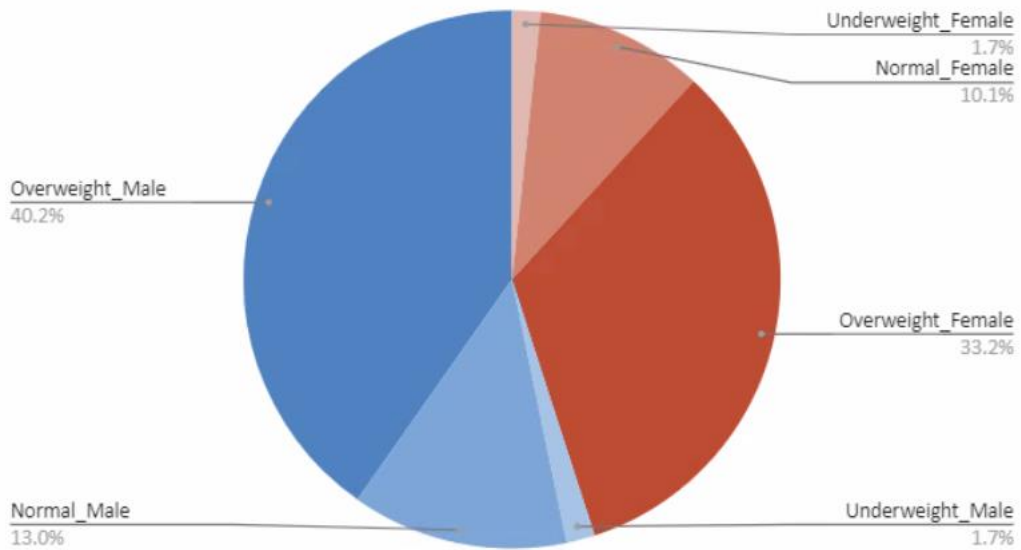Overweight_Female 33.2%
Normal_Male 13.0%
Underweight_Male 1.7%

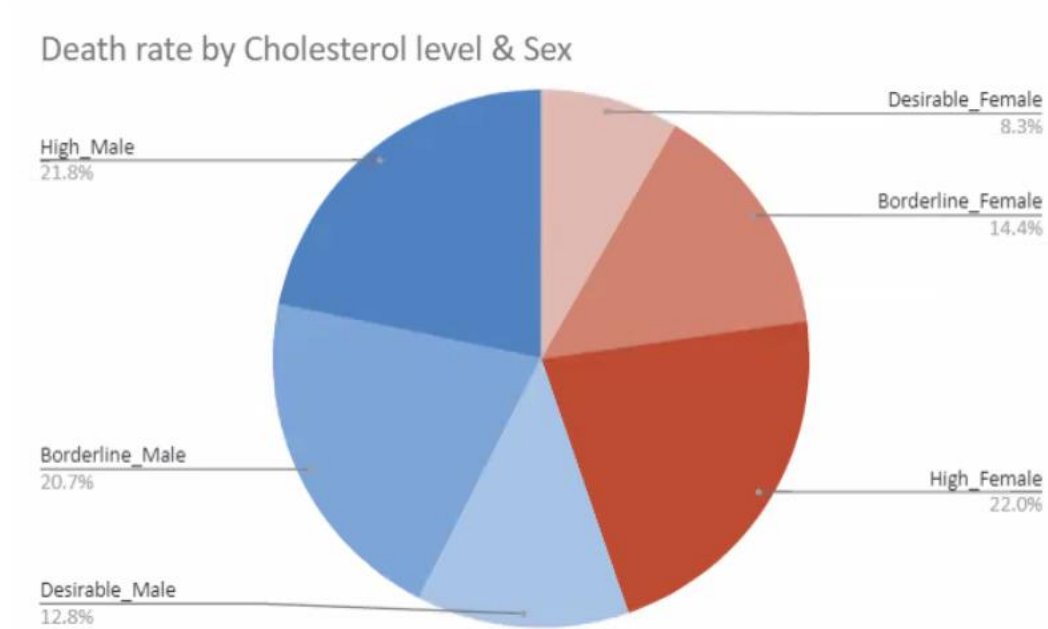Figure 7- Death rate by weight and Sex

Figure 8- Death rate by Cholesterol level and Sex

In Figure 6 Death rate by blood pressure level and sex is female 27.8% (high) and male 31.1% (high). In Figure 7 Death rate by weight and sex is female 33.2% (overweight) and male 40.2% (overweight). In Figure 8 Death by cholesterol level and sex is female 22% (high) and male 21.8% (high).

RECOMMENDATIONS

I would recommend an educational program that teaches the importance of a healthy lifestyle. These programs would include exercise programs, ways to create nutritional meals, limit tobacco products, etc.

I would recommend adopting a healthy lifestyle and this would include creating goals and optimize sleep patterns. As Ill as advocate routine medical check-ups.

I would recommend continued research on cardiovascular diseases which I believe can potentially improve prevention measures and have a more in-depth understanding of how to diagnose and treat the diseases.

I recommend ways to manage stress through teaching coping mechanisms.